DOCUMENT RESUME

ED 128 376 TM 005 498

AUTHOR	Sherman, Robert E.; And Others
TITLE	Program Evaluation Project Report, 1969-1973. Chapter Four: An Examination of the Reliability of the Kiresuk-Sherman Goal Attainment Score by Means of Components of Variance.
INSTITUTION	Program Evaluation Resource Center, Minneapolis, Minn.
SPONS AGENCY	National Inst. of Mental Health (DHEW), Rockville, Md. Div. of Mental Health Services Program.
PUB DATE	Aug 74
GRANT	NIMH-5-R01-1678S04
NOTE	15p.; For related documents, see TM 005 495-501
AVAILABLE FROM	Program Evaluation Project, 501 Park Ave. South, Minneapolis, Minnesota 55415 ($1.00)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Analysis of Variance; Evaluation Methods; *Goal Orientation; Interviews; Measurement Techniques; *Mental Health Programs; *Program Evaluation; *Reliability; Scores; *Statistical Analysis
IDENTIFIERS	*Goal Attainment Scaling

ABSTRACT

The P.E.P. Report 1969-1973 focuses on the various findings and activities of the Program Evaluation Project. The study in this chapter was designed to conduct a statistical analysis of the Goal Attainment Score, and estimate variance components due to choice of material in the followup guide, followup interviewer bias or error, and the client's actual long-term deviation from expectation. These factors together determine the reliability of the Goal Attainment score as it was applied in this Program Evaluation Project study, and, in addition, provide some useful indication of its potential reliability in other evaluative applications. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128376

CHAPTER FOUR

AN EXAMINATION OF THE RELIABILITY OF THE KIRESUK-SHERMAN GOAL ATTAINMENT SCORE BY MEANS OF COMPONENTS OF VARIANCE.

P.E.P. REPORT 1969 - 1973

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

A REPORT ON FOUR YEARS OF STAFF EFFORT AT THE PROGRAM EVALUATION PROJECT.

TM005 498

CHAPTER FOUR

Program Evaluation Project Report, 1969-1973

AN EXAMINATION OF THE RELIABILITY OF THE KIRESUK-SHERMAN GOAL ATTAINMENT SCORE BY MEANS OF COMPONENTS OF VARIANCE

Prepared by:

Robert E. Sherman, Ph.D.
James W. Baxter
Donna M. Audette

August, 1974

Thomas J. Kiresuk, Ph.D., Director
Program Evaluation Project
501 Park Avenue South
Minneapolis, Minnesota 55415

Developed under Grant #5 R01 1678904, National Institute of Mental Health, Department of Health, Education, and Welfare

---

Acknowledgements: Thanks to all of the Mental Health Service staff who participated in the construction of the Goal Attainment Follow-up Guides for this study, and to all of the Program Evaluation Project staff, in particular, William G. Makela, who was instrumental in the operationalizing of the study in 1970.

TABLE OF CONTENTS

	PAGE NUMBER
General Introduction to the P.E.P. Report 1969-1973	1
Synopsis	2
I. Introduction	3
A. Goal Attainment Scaling Methodology, General	3
B. Goal Attainment Scaling Methodology, As Used at Hennepin County Mental Health Service	3
C. The Relationship of Reliability to Validity for the Goal Attainment Score	4
II. Study Objectives and Design	4
III. Results	4
A. Course of Study	4
B. The Model for Analysis	5
C. Variance Component Estimates	6
D. Reliability Coefficients	7
IV. Conclusions and Summary	7
A. Clients Making Their Own Follow-up Guides	8
B. Negotiating the Follow-up Guide with the Client	8
C. Multiple Follow-ups	8
D. Therapists Conducting Their Own Follow-ups	8
E. Semi-Standardized Scales	8
F. The Goal Attainment Process as a Part of Therapy	9
Program Evaluation Project Staff Listing	10

For further information, please contact Ms. Joan Brintnall, Program Evaluation Project, 501 Park Avenue South, Minneapolis, Minnesota 55415.

GENERAL INTRODUCTION TO THE P.E.P. REPORT 1969-1973

The P.E.P. Report 1969-1973 focuses on the various findings and activities of the Program Evaluation Project. It is being published in pamphlet form with one pamphlet for each chapter.

As of January, 1974, the Program Evaluation Project, whose title was changed to the Program Evaluation Resource Center as of June, 1974, is funded by a three year collaborative grant with the Mental Health Services Division of the National Institute of Mental Health. The purpose of the grant is to emphasize the coordination and dissemination of information on a variety of program evaluation methodologies, especially Goal Attainment Scaling.

Further information on the Goal Attainment Scaling methodology and program evaluation is available in other written and recorded materials from the Program Evaluation Resource Center office. At this time various other chapters of the P.E.P Report 1969-1973 are available, including Chapter One, "Basic Goal Attainment Scaling Procedures", Chapter Two, "Activities of the Follow-up Unit", Chapter Three. "An Introduction to Reliability and the Goal Attainment Scaling Methodology", Chapter Five, "A Construct Validity Overview of Goal Attainment Scaling" and Chapter Nine, "Evaluation of the Adult Outpatient Program, Hennepin County Mental Health Service". Additional chapters will be released this year as they are completed.

SYNOPSIS FOR CHAPTER FOUR

AN EXAMINATION OF THE RELIABILITY OF THE KIRESUK-SHERMAN

GOAL ATTAINMENT SCORE BY MEANS OF COMPONENTS OF VARIANCE

PURPOSE: The study in this chapter was designed to conduct a statistical analysis of the Goal Attainment Score, and estimate variance components due to choice of material in the follow-up guide, follow-up interviewer bias or error, and the client's actual long-term deviation from expectation. These factors together determine the reliability of the Goal Attainment score as it was applied in this Program Evaluaiton Project study, and, in addition, provide some useful indication of its potential reliability in other evaluative applications.

MAJOR FINDINGS: Two Goal Attainment Follow-up Guides were independently completed on each of 44 clients. Each client was followed-up twice by different follow-up interviewers, and each follow-up guide scored on each occasion. Thus, each client yielded four Goal Attainment scores. Analyzing these data by a components of variance model yielded estimated score variances of 47.70 (50%) due to client long-term deviation from expectation, 14.53 (15%) due to short-term client changes or follow-up bias fluctuations, 16.12 (17%) due to choice of follow-up guide material, and 17.93 (18%) due to follow-up interviewer errors in scoring or observation.

These findings are then related to various suggested modifications in the Goal Attainment Scaling procedure.

I. INTRODUCTION

The purpose of the study on Goal Attainment Scaling by the Program Evaluation Project staff was to examine the feasibility of shifting the emphasis in program evaluation away from process factors (such as volume, load, etc.) toward measures of outcome reflecting attainment of individualized clinical goals (alleviation of depression, vocational adjustment, etc.). This report presents a detailed discussion of one reliability study of the Goal Attainment Scaling methodology utilized at the Hennepin County Mental Health Service.

A. Goal Attainment Scaling Methodology, General

The Goal Attainment Scaling methodology is a client-specific method of goal setting and evaluation. The methodology allows the goal setter to establish unique goals and levels of attainment for individual clients while retaining the ability to make outcome comparisons. Its basic characteristics are: 1) establishing a set of specific goals with or for the client; 2) assigning weights (w_i) to each goal relative to its outcome significance; 3) projecting a follow-up date; and 4) establishing a well-defined set of attainment levels for each projected goal. At the prespecified follow-up date the levels of attainment (x_i) on all specified goals are determined. These attainment levels, given values from -2 to +2, and the relative goal weights (any set of positive values), are used to generate a standardized Kiresuk-Snerman "Goal Attainment score", Y.

$$Y = 50 + \frac{10\Sigma w_i x_i}{\sqrt{(1-\rho)\Sigma w_i^2 + \rho(\Sigma w_i)^2}},$$

where ρ is taken to be .3.

B. Goal Attainment Scaling Methodology, As Used At Hennepin County Mental Health Service

In the application of Goal Attainment Scaling at the Hennepin County Mental Health Service, follow-up guides were constructed for all new clients during the intake process. This intake process consisted of one or two diagnostic interviews, usually included completion of psychological testing and, when necessary, a medication consultation. It was the intake clinician's responsibility to complete a follow-up guide with a minimum of three goals for each intake case. A typical follow-up guide constructed for use in the research study is shown in Figure I.

Care was taken to insure the "follow-up-ability" of the goals on the follow-up guides. The follow-up guides were reviewed by members of the research staff for problems which might interfere with the scoring of the follow-up guides. Problems were negotiated with the follow-up guide constructor for clarification or change. Clients were then assigned to a treatment mode. The assignment was random, if ethically possible.

FIGURE I: Sample Goal Attainment Follow-up Guide

GOAL ATTAINMENT FOLLOW-UP GUIDE

Levels of Attainment	Scale 1 Wt. = 2 Family Communication	Scale 2 Wt. = 3 Admitting problem	Scale 3 Wt. = 1 Education	Scale 4 Wt. = 2 Living Arrangements
Most Unfavorable Outcome Thought Likely with Therapy	I refuse to stay in same room with my parents at all. Leave room immediately.	I can't admit that I have difficulties to anyone except myself.	No desires and no plans to go back to school.	I live alone in an apartment or single room.
Less Than Expected Success with Therapy	Will not stay in same room with parents for more than 10 mintues.	I can admit that I have some physical problems, but no emotional problems.	I want to go back to school, but have no specific plans for doing so.	I live with my parents.
Expected Level of Success with Therapy	Will stay in same room with parents for 11 to 20 mintues.	I can admit only one or two days per week that I have one emotional problem.	I want to go back to school, and have made plans (collected information on school, thought about courses).	I live with relatives other than my parents.
More Than Expected Success with Therapy	Will stay in same room with parents for more than 20 minutes, but only if someone else is present.		I want to go to school and have made plans and have enrolled in one course.	I live with one or more non-relatives but don't have close friendships with any of them.
Most Favorable Outcome Thought Likely with Therapy	Will stay in same room with parents for more than 20 minutes, even if no one else is around.	I can admit almost any day that I have more than one emotional problem.	As above, and have enrolled in more than one course.	I live with non-relatives and have a close friendship with at least one of them.

At the specified follow-up time, "moonlighting" social workers from other local agencies would personally interview the client and score the follow-up guide. These scores were withheld from the Mental Health Service staff until the conclusion of the study.

C. The Relationship of Reliability to Validity for the Goal Attainment Score

Under suitable assumptions, Sherman (1974) has observed that the validity of the Goal Attainment score can be established through the content validity argument. This argument concludes that the Goal Attainment score, by its nature and by what it is represented to measure, is as valid as it is reliable. This conclusion emphasizes the importance of a detailed examination of the Goal Attainment score reliability.

II. Study Objectives and Design

A satisfactory appraisal of the reliability of the Goal Attainment score must address at least the following questions:

a. What is the total amount of variation of Goal Attainment scores in the measured population?

b. How much of the total variation is due to the particular Goal Attainment Follow-up Guide that happened to have been made for a client (i.e., if a client had seen a different intake interviewer, an altogether different Goal Attainment Follow-up Guide might have been made)?

c. How much of the total variation is due to observation or scoring errors in follow-up?

d. How much of the total variation is due to the particular moment of the follow-up interview? (In our case, follow-up interviews were made about six months after assignment to treatment; one would hope that choosing five or seven months instead, would have little effect on the outcome measure.)

e. Finally, how much of the total variation can be assigned to the client, independent of the particular Goal Attainment Follow-up Guide, follow-up time, and observation error? The element creating this variation is what we are trying to measure.

To answer these questions efficiently, an analysis of variance model was chosen that required two follow-up guides on each subject, and two follow-ups on each follow-up guide. Thus, each subject would yield four Goal Attainment scores, one from each follow-up guide on each follow-up interview. It was judged that sufficient accuracy could be achieved with 40 subjects.

All adult outpatients of the Mental Health Service would have follow-up guides constructed for them during the intake process. The second follow-up guide required for the reliability study would be obtained from the assigned therapist. The therapist would tailor his follow-up guide to the follow-up date specified by the intake interviewer (usually six months to a year after treatment assignment) but would be otherwise unaware of the material on the intake interviewer's follow-up guide.

To insure that each follow-up guide received about equal attention in the follow-up interview, and to minimize the likelihood of a follow-up interviewer recognizing the follow-up guide origin from its content, the scales from the two follow-up guides were randomly mixed and typed on a single master follow-up guide. (The scales were separated later for the analysis.)

At approximately the prespecified follow-up date, the master guide would be scored simultaneously in a follow-up interview and then scored again in another follow-up interview (by a different interviewer) about two weeks later.

III. Results

A. Course of the Study

From May 1970 to October 1972, dual follow-up guides were completed on 84 clients. Of these, 44 were successfully followed-up twice. The reasons for the failures were: 17 clients were unlocatable for either the first or second follow-up interview; 15 clients refused to participate in either the first or second follow-up interview; and for eight clients, other criteria were not met, such as poor follow-up guide construction on clients not having completed the minimum of two therapy sessions in their assigned mode prior to the prescribed follow-up date.

Of the 44 successfully followed-up subjects, 29 (66%) were female, and ages ranged from 18 to 52, with an average age of 27. These and other client characteristics are similar to those of the rest of the Mental Health Service client population. (More detail can be found in chapter six of the P.E.P. Report, 1969-1973.)

Subjects were treated by Individual Therapy (33, or 75%); Group Therapy (6, or 14%); Marriage Counseling (3, or 7%); Day Care Treatment (1, or 2%); and Medication Clinic (1, or 2%). The professions of the Mental Health Service staff were represented in both the in-

take interview and therapy functions. Most were social workers though psychiatrists, psychologists and psychiatric nurses also participated in approximate proportion to their numbers on the Mental Health Service staff.

The length of time between the first and second follow-ups ranged from 5 to 67 days, with a mean of 25 days (see Table I). To investigate the effect of time between follow-ups on the size of the difference between Goal Attainment scores from the two follow-up times, all clients' differences in average Goal Attainment scores at first and second follow-ups (absolute values) were ranked; times between follow-ups were ranked; and a Spearman rank order correlation coefficient was computed. The value was $r_s = .12$ ($N = 44$), far from significance.

The Goal Attainment score on either follow-up guide from either follow-up had means and standard deviations close to the expected values of 50 and 10, respectively (see Table I). Table I also gives the means for the sample total, as well as means for a breakdown of the sample by the number of days between follow-ups.

TABLE I: Mean Goal Attainment Scores for Both Follow-up Interviews and Both Follow-up Guides by Number of Days Between Follow-up Interviews

Number of days between first and second interview:	5-14	15-29	30-44	45-67	TOTAL
Number of subjects:	N = 12	N = 17	N = 11	N = 4	N = 44
FIRST INTERVIEW					
Intake Interviewer G.A.S.	45.55	50.97	47.61	50.62	48.62 S.D.= 9.18
Therapist G.A.S.	48.13	55.56	47.53	54.77	51.45 S.D.= 9.84
SECOND INTERVIEW					
Intake Interviewer G.A.S.	44.68	51.87	51.99	50.63	49.83 S.D.=11.18
Therapist G.A.S.	49.04	55.68	54.91	54.57	53.57 S.D.= 8.89

TABLE II: Counts of Clients by First and Second Follow-up Interviewers

SECOND FOLLOW-UP INTERVIEW

Interviewer Code	A	B	C	D	E	F	G	Total Follow-ups	Intake Interviewer Mean G.A.S.
A		1		1		4	1	7	48.83
B				1	1	1	1	4	46.89
C	1			1		1		3	48.17
D		3	3		2	1		9	50.01
E		4		1			1	6	47.33
F	1	6		2			3	12	47.01
G		1		1		1		3	55.73
Total Follow-ups	2	15	3	7	3	8	6	44	48.62
Intake Intv. Mean G.A.S.	60.09	46.38	59.57	46.37	47.87	50.95	53.63	49.82	

All follow-up interviews were conducted by master's level social workers. In no case were both follow-up interviews conducted by the same interviewer, and though a random assignment of follow-up interviewers was not implemented, an attempt was made to avoid consistent linkages between first and second follow-up interviewers (see Table II). Simple analyses of variance did not show statistically significant differences in average scores by follow-up interviewers.

B. The Model for Analysis

In order to use analysis of variance methods to identify variance components for the Goal Attainment score, it is necessary to specify a detailed statistical model:

Let Y_{ijk} represent the Goal Attainment score from the kth follow-up on the jth follow-up guide on the ith patient. We then define the model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk},$$

where i goes from 1 to I (I=44), j goes from 1 to J (J=2), and k goes from 1 to K (K=2), and we assume

μ is a true mean effect,

α_i are random effects representing the ith client's true long-term average deviation from μ, and the α_i are NID (Normally and Independently Distributed) $(0, \sigma_\alpha^2)$.

β_j are fixed effects representing the different sources of follow-up guides (the first one created by the intake worker, or the second one created by the therapist), and $\sum_j \beta_j = 0$.

γ_k are fixed effects representing the effect of the follow-up order, that is, a combination of experience effect and true average client change across time from first to second follow-up, and $\sum_k \gamma_k = 0$.

$(\alpha\beta)_{ij}$ are random effects due to the jth guide on the ith client, and represents a deviation from a conceptual average score of an infinite number of independently created follow-up guides on the same individual, and the $(\alpha\beta)_{ij}$ are NID $(0, \sigma_{\alpha\beta}^2)$.

$(\alpha\gamma)_{ik}$ are random effects due to either true fluctuations in the state of the client from time to time, or fluctuations in the "optimism" of the follow-up interviewers from time to time, and the $(\alpha\gamma)_{ik}$ are NID $(0, \sigma_{\alpha\gamma}^2)$.

$(\beta\gamma)_{jk}$ are fixed effects due to the interaction of follow-up guide source and follow-up time. That is, the "learning effect", or true average client change across time may be different for follow-up guides from different sources; and $\sum_j (\beta\gamma)_{jk} = \sum_k (\beta\gamma)_{jk} = 0$.

ε_{ijk} are residual random errors of observation or scoring, and the ε_{ijk} are NID $(0, \sigma_\varepsilon^2)$.

The task is now to analyze the observed scores in terms of the above parameters, estimating the size and testing the significance of the estimated variance components.

Though the analysis of variance which follows at first appears to be based on a three-factor factorial design with one random and two fixed effects (and in fact the sum of squares is broken down in that fashion), the expected mean squares do not conform to that model. Because of the assumption that the $(\alpha\beta)$ and $(\alpha\gamma)$ "interactions" were random variables, the design has characteristics of a "nested" or hierarchical design.

The usual F-ratio tests demonstrate statistical significance at the .01 level for the effects of "Individuals", "Source of Guide", "Individual x Source" interaction, and "Individual x Follow-up Order" interaction.

TABLE III:

Analysis of Variance

44 Subjects, Each With Four Goal Attainment Scores Generated According to the Reliability Study Model

SOURCE OF VARIATION	df	MS	E(MS)
Individuals	I-1=43	270.04**	$\sigma_\varepsilon^2 + 2\sigma_{\alpha\beta}^2 + 2\sigma_{\alpha\gamma}^2 + 4\sigma_\alpha^2$
Source of Guide	J-1=1	476.03**	$\sigma_\varepsilon^2 + 2\sigma_{\alpha\beta}^2 + 88\sum_j \beta_j^2$
Follow-up Order	K-1=1	121.66	$\sigma_\varepsilon^2 + 2\sigma_{\alpha\gamma}^2 + 88\sum_k \gamma_k^2$
Indiv. x Source	(I-1)(J-1)=43	50.17**	$\sigma_\varepsilon^2 + 2\sigma_{\alpha\beta}^2$
Indiv. x F.U. Order	(I-1)(K-1)=43	46.99**	$\sigma_\varepsilon^2 + 2\sigma_{\alpha\gamma}^2$
Source x F.U. Order	(J-1)(K-1)=1	9.17	$\sigma_\varepsilon^2 + 44\sum_{jk}(\beta\gamma)_{jk}^2$
Residual F.U. Error	(I-1)(J-1)(K-1)=43	17.93	σ_ε^2

**Significant at the p<.01 level

C. Variance Component Estimates

Using the analysis of variance table, the variance components together with 90 percent confidence limits on the estimates may now be computed. (Scheffé, 1959)

σ_ε^2, the residual error variance due to errors of observation or scoring in follow-up, is estimated by $s_\varepsilon^2 = 17.93$, with 90 percent confidence interval 13.00 to 26.58. That is, we might expect a random error with a standard deviation of about four points in the Goal Attainment score due to the follow-up interviewer's errors of observation or scoring.

$\sigma_{\alpha\beta}^2$, the error variance due to the construction of the Goal Attainment Follow-up Guide and the material chosen for inclusion is estimated by

$$s_{\alpha\beta}^2 = (50.17 - 17.93)/2 = 16.12,$$

with 90 percent confidence interval 8.10 to 28.65. That is, we might expect a random error with a standard deviation of about four points (the square root of 16.12) in the Goal Attainment score due to the material chosen for the follow-up guide. This is the error component unique to the Goal Attainment Scaling procedure. A standardized "fixed" test would have no such component, but such "fixed" tests could be less relevant to a particular client's problems.

$\sigma_{\alpha\gamma}^2$ the variance component due to fluctuations over time in either the true state of the client, or the general optimism of the follow-up interviewers, is estimated by

$$s_{\alpha\gamma}^2 = (46.99 - 17.93)/2 = 14.53,$$

with 90 percent confidence interval 6.86 to 26.25. To the extent that $\sigma_{\alpha\gamma}^2$ is due to the true state of the client at the follow-up time, we may not wish to consider it an "error". While a measure which would give the long-term average status of a client rather than his exact condition at a particular moment might be preferred, such a measure cannot be approached without repeated observations across time. It should, therefore, not stand against a one-time measure if it only measures the status of a client at the time of the measurement. But this variance component may also be due to variations in the level of optimism of the follow-up interviewer. That is, how generous is the follow-up interviewer in his interpretation of the client behavior. In this case $\sigma_{\alpha\gamma}^2$ would be an error variance.

σ_α^2, the variance component due to differences among clients in their true long-term average deviation from expectation is estimated by:

$$s_\alpha^2 = (270.04 - 50.17 - 46.99 + 17.93)/4 = 47.70,$$

with 90 percent confidence limits 31.16 to 71.21. That is, if all measurement errors could be excluded, we would be left with a Goal Attainment score standard deviation of about seven, instead of the 10 which is observed.

D. Reliability Coefficients

In its intended application, the Goal Attainment score is computed from a single follow-up on a single Goal Attainment Follow-up Guide. Thus, in the model for the score, Y_{ijk}, the j and k are always 1, and components that vary only with j or k are now constant across all observations and absorbed into the "true mean effect", μ. The model then becomes:

$$Y_i = \mu + \alpha_i + (\alpha\gamma)_i + (\alpha\beta)_i + \varepsilon_i,$$

where the components represent the same effects as before, but now varying only across.

The variance of Y_i is then constructed as follows:

$$\sigma_y^2 = \sigma_\alpha^2 + \sigma_{\alpha\gamma}^2 + \sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2$$

which may be estimated by:

$$s_y^2 = s_\alpha^2 + s_{\alpha\gamma}^2 + s_{\alpha\beta}^2 + s_\varepsilon^2$$

$$= 47.70 + 14.53 + 16.12 + 17.93$$

$$= 96.28$$

for which a 90 percent confidence interval may be computed to be 79.14 to 113.41.

The components of variance can be related to the total variance of a Goal Attainment score (see Figure II), and we may respond to the questions posed in Section I, item C, viz.,

a. What is the total amount of variation of the Goal Attainment scores in the measured population?

Answer: The variance of the score is estimated at 96.28, or a standard deviation of 9.81.

b. How much of this total variation is due to the particular Goal Attainment Follow-up Guide that happened to have been made for each client?

Answer: The variance component due to the choice of guide material is estimated at 16.12, or 17 percent of the total score variance.

c. How much of total variation is due to errors of observation or scoring?

Answer: The variance component due to follow-up error is estimated at 17.93, or 18 percent of total score variance.

d. How much of the total variation is due to the particular moment of follow-up?

Answer: Here the experimental design could not separate short term client fluctuations from follow-up interviewer bias. These two components together contribute an estimated variance component of 14.53, or 15 percent of the total score variance.

e. How much of the total variation can be assigned to the client, independent of the particular Goal Attainment Follow-up Guide, follow-up time, and observation error?

Answer: The variance component assignable to differences among clients in their long-term deviation from expectation is estimated by 47.70, or 50 percent of the total score variance.

The above information can be expressed in terms of various reliability coefficients, viz.

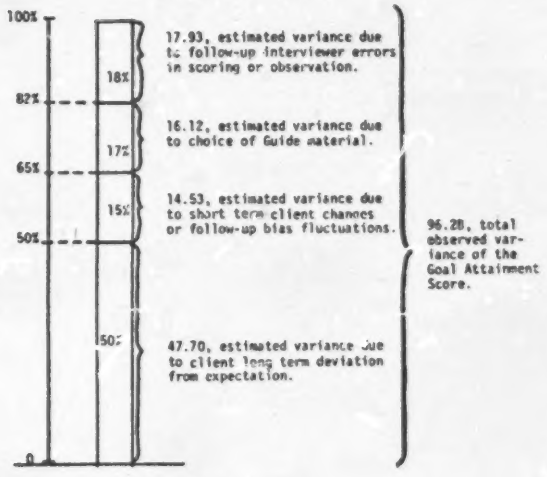
How well does the Goal Attainment score reflect the long-term status of the client?

We estimate:

$$r_1 = \frac{s_\alpha^2}{s_y^2} = \frac{47.70}{96.28} = .50$$

FIGURE II

BREAKDOWN OF VARIANCE COMPONENTS OF THE GOAL ATTAINMENT SCORE


100%
82%
65%
50%
0
18%
17%
15%
50%
17.93, estimated variance due to follow-up interviewer errors in scoring or observation.
16.12, estimated variance due to choice of Guide material.
14.53, estimated variance due to short term client changes or follow-up bias fluctuations.
47.70, estimated variance due to client long term deviation from expectation.
96.28, total observed variance of the Goal Attainment Score.

Or, how well does the Goal Attainment score reflect the actual status of the client at the time of follow-up? Here again is the problem of question four, above. How much of $s_{\alpha\gamma}^2$ can we assign to the client status (which we wish to measure) and how much to extraneous interviewer bias? Depending upon this division, we estimate the reliability of the Goal Attainment score to be:

$$\frac{s_\alpha^2}{s_y^2} = .50 \leq r_2 \leq .65 = \frac{s_\alpha^2 + s_{\alpha\gamma}^2}{s_y^2}$$

Similarly, we can bracket the reliability of follow-up scoring:

$$\frac{s_\alpha^2 + s_{\alpha\beta}^2}{s_y^2} \quad .66 \leq r_3 \leq .81 = \frac{s_\alpha^2 + s_{\alpha\beta}^2 + s_{\alpha\gamma}^2}{s_y^2}$$

And, finally, the reliability of follow-up guide construction when the constructors compared are intake interviewers and therapists is estimated to be:

$$r_4 = \frac{s_\alpha^2 + s_{\alpha\gamma}^2 + s_\varepsilon^2}{s_y^2} = .83$$

It should be emphasized here that it is r_1 or r_2 that reflect the reliability of the Goal Attainment score in its application. The coefficients r_3 and r_4 might be considered "special interest" statistics.

IV. Conclusions and Summary

It is now clear that the Goal Attainment

score measured at least the degree to which a client's outcome status (on plausibly mental health related characteristics) conformed to the expectations of mental health professionals. The most complete picture of the score reliability is obtained by examining the variance component estimates presented in the previous section. From these, two "reliability coefficients" were computed as candidates to represent the Goal Attainment score reliability, r_1 (= .50), and r_2 (between .50 and .65). It simplifies the statement of this result to use an average figure of r = .57 to represent the reliability of the Goal Attainment Scaling application used in the Program Evaluation Project study. Clearly, more refined analysis of our data would not greatly change this estimate.

Is Goal Attainment Scaling ready for practical evaluative applications? The most critical point in the process is surely follow-up guide construction. Without thoughtfully and skillfully constructed follow-up guides, both follow-up guide construction and follow-up determination errors may become too large. Even with considerable care (in both follow-up guide construction and follow-up) the reported reliability of .57 is only moderately high, though it does take into account all the errors encountered in the application. That is, both follow-up determination errors (which includes both test-retest and inter-rater differences) are accounted for in the reported r of .57. (Some reported reliability coefficients are either "alternate form" or "test-retest" reliability, but not both, and therefore may not represent the practical reliability of a score.) Given the severity of our test and the unique advantage of the Goal Attainment Scaling technique (i.e., completely individualized goals), the authors consider the Goal Attainment score acceptably reliable in the Program Evaluation Project application.

However, the Program Evaluation Project application is basically research-oriented. Most evaluators face significantly different circumstances, programs, and overall objectives for the evaluation process. There may not be sufficient staff to permit independent follow-up guide construction and follow-up interviews, or it may be desired that the client set his own goals. Improvement of outcome rather than the evaluation of therapy may be the immediate objective and, of course, a high cost evaluation program may be difficult to justify. There have been several attempts to modify the Goal Attainment Scaling procedure to make it more compatible with one or more such specifications. Though work is still in progress, it is useful to briefly consider, in light of this study, the reliability implication of some of the suggested procedure modifications.

A. Clients Making Their Own Follow-up Guides

If all clients were to make their own follow-up guides, it could save staff time, remove therapist bias from the follow-up guide content, greatly improve follow-up guide construction reliability, and could also reduce errors of determination in the follow-up (the client should know what he meant when he specified the scales). A step-by-step manual for the client to use in doing this has been developed (Garwick, 1973). The chief disadvantage of this modification is that the client may lack the skill or insight to determine realistic goals and attainment levels.

B. Negotiating the Follow-up Guide With the Client

If the therapist were to negotiate the Goal Attainment Follow-up Guide with the client, we might hope to obtain many of the benefits of the client making the follow-up guide himself (as suggested above) while eliminating through the negotiations many of the inappropriate or unrealistic goals or attainment levels. This has been suggested by Sherman (1972) and applied by Lombillo, et. al. (1973). A related benefit of this modification is that good concrete communication between therapist and client with respect to therapy goals is necessarily established in the beginning. The chief disadvantage is that therapists may be suspected of developing a self-serving approach to the negotiation.

C. Multiple Follow-ups

Multiple follow-ups on Goal Attainment Follow-up Guides has been suggested as a way of following either the course of therapy or the durability of therapy results. Multiple follow-ups would also permit the reduction of follow-up determination error, and the smoothing of short-term client status fluctuations. Its chief difficulty is cost, along with the fact that clients may tire of cooperating, or be unlocatable.

D. Therapists Conducting Their Own Follow-ups

If the therapist were to conduct the follow-up, he would have the advantage of his clinical experience with the client to assist in the interpretation of the client's behavior, and follow-up determination error should be reduced. Feedback would be immediate. He could use his acquired rapport to conduct inexpensive follow-up interviews by phone, making multiple follow-ups more practical. This modification suffers the possibility of therapist bias.

E. Semi-Standardized Scales

It could simplify the construction of the Goal Attainment Follow-up Guide and provide an easier starting point for categorizing clients

by follow-up guide content, if goals were selected from some finite list, perhaps each with a well-constructed set of graded attainment levels to choose from. This might also reduce follow-up guide construction variance, and follow-up determination error as well. Its major disadvantage is that follow-up guides may be less relevant to the client's specific problems.

F. The Goal Attainment Process as a Part of Therapy

It has been suggested that the goal setting process is itself a useful part of therapy. In this model, reliability may be of little concern.

Many of the modifications in the Goal Attainment Scaling procedure mentioned above are being attempted. While the results are not yet in, it does appear that Goal Attainment Scaling is moving successfully from research to practical evaluative applications.

References.

Garwick, G. Client characteristics for three adult outpatient groups. P.E.P. Report 1969-1973. Chapter Six.

Garwick, G. Guide to goals I. Unpublished Program Evaluation Project report, 1973.

Kiresuk, T.J. & Sherman, R.E. Goal Attainment scaling: a general method for evaluating comprehensive community mental health programs, Community Mental Health Journal, 1968, 4(6), 443-453.

Lombillo, J., Kiresuk, T.J., Sherman, R.E. Contract fulfillment analysis: evaluating a community mental health program: Hospital & Community Psychiatry, November, 1973, Volume 24, Number 11, 760-762.

Sheffé, H. The analysis of variance. New York: John Wiley and Sons, Inc., 1959. 231-235

Sherman, R.E. Contract fulfillment scaling. Unpublished Program Evaluation Project Report, 1972.

Sherman, R.E. Position paper on validity. Unpublished Program Evaluation Project Report, 1974.

PROGRAM EVALUATION PROJECT STAFF LISTING

CURRENT STAFF MEMBERS

Thomas J. Kiresuk, Ph.D.
Principal Investigator 1969-1974

James Baxter
Research Assistant 1970
Operations Manager 1971
Assistant Coordinator 1972-1973
Re-design Coordinator 1974

Diane Berg
Editorial Secretary 1973-1974

David Bolin
Assistant Editor 1974

Joan Brintnall
Student Assistant 1971
Secretary-Receptionist 1972
Administrative Assistant 1973
Dissemination, Consultation,
and Utilization Supervisor 1974

Joan Dreyer
Research Clerk 1972
Research Assistant 1973
Research/Administrative Clerk 1974

Geoffrey Garwick, M.A.
Research Applications Consultant 1970
Program Evaluation Coordinator 1971
Deputy Assistant Director 1972
Assistant Director 1973
Dissemination, Consultation,
and Utilization Consultant 1974

Carolyn Jasperson
Student Assistant 1973-1974

Laurence Kivens, M.A.
Applications Analyst 1971
Editorial Supervisor 1971
Editor 1972-1974

Mary Knepper
Appointment Interviewer 1970-1973
Administrative Assistant 1974

Judy Long
Administrative Secretary 1974
Research Assistant 1974

Sander Lund
Management Applications Supervisor 1971
Assistant Coordinator 1971-1972
Coordinator for Administration 1973
Assistant Director 1974

Nancy Petersen
Secretary 1973
Research Assistant 1974
Follow-up Supervisor 1974

Michael Saunders
Programmer Analyst 1971-1972
Programmer Supervisor 1973-1974

Robert Sherman, Ph.D.
Associate Investigator 1969-1974

Vicki Stoleson
Secretary-Receptionist 1974

Mary Ellen Whalen
Student Assistant 1973-1974
Research Analyst 1974

PREVIOUS STAFF MEMBERS

Donna M. Audette
Research Assistant 1970
Follow-up Assistant 1971
Follow-up Supervisor 1972-1973
Utilization Consultant 1974

Janis Bibee, M.A.
Editorial Secretary 1972
Editorial Assistant 1973
Assistant to the Editor 1974

Anita Bjornson
Research Assistant 1970

Barbara Blazick
Student Assistant 1970

Mary Duroche
Editorial Secretary 1972

David Feigal
Research Assistant 1970-1971

Thomas Griffin
Student Assistant 1970

Marilee Grygelko
Student Assistant 1971-1972
Research Assistant 1973

Edward Gubman
Student Assistant 1971

Colleen Halley
Student Assistant 1971-1973

Susan Jones
Research Assistant 1971
Research Associate 1972-1974

Robert Kearney
Editorial Assistant 1972

Karen Kohout, M.A.
Research Analyst 1969
Research Supervisor 1970-1971

Sherry Lampman
Administrative Assistant 1970
Linguistic Analysis Consultant 1971
Content Analysis Supervisor 1973

William Makela, M.A.
Follow-up Supervisor 1970-1972

Charles Meade
Research Assistant 1970-1972

Deirdre Meade
Secretary-Receptionist 1971
Administrative Assistant 1972

Sylvia Muilenberg
Administrative Assistant 1969
Administrative Supervisor 1970

Nils Olsson
Student Assistant 1971
Research Assistant 1972-1973

Carol Pollock
Research Clerk 1972

William Prock
Community Applications Supervisor 1971-1972

Peter Ree
Student Assistant 1970

Martha Rosen
Secretary-Receptionist 1970
Editorial Assistant 1971-1972

Susan Salasin
Coordinator 1969-1970
Assistant Director 1971-1972
Research Applications Consultant 1973

Richard Tripp
Programming Supervisor 1970
Design and Analysis Supervisor 1971-1972

Mary Trone
Editorial Secretary 1973

Roger Twedt
Student Assistant 1974

Carol Vanderpool
Secretary-Receptionist 1971-1972
Administrative Assistant 1973

Cynthia Wetterland
Secretary-Receptionist 1970

Allen Wichelman
Medical Records Clerk 1970
Management Applications Supervisor 1971

Sue Wright
Clinical Applications 1971

Carole Zimbrolt
Research Analyst 1969-1971

FOLLOW-UP INTERVIEWERS - CURRENT

Kathleen Bergum, M.S.W.

Charles Besnett, M.S.W.

Carol Dethmers, B.A.

Marcia Frankenberg, B.A.

George Meirick, M.S.W.

FOLLOW-UP INTERVIEWERS - PREVIOUS

Mary Ann Anzelc, R.N.

James Bergum, M.S.W.

Roanne Borkon, R.N.

Larry Bultena, M.S.W.

Scott Craven

Jeanine Emmons, R.N.

Barbara Gusek, R.N.

Mary Keturakat, R.N.

Steve Lapinsky

Betty Metz, B.A.

Madiline Sachs, R.N.

James Snope, M.S.W.

Milt Somerfleck, M.S.W.

CONSULTANTS

Dean Beaulieu, Ph.D.
Clinical Coordinator 1971-1974

James Boen, Ph.D.
Statistical Consultant 1969-1973

Byron Brown, Ph.D.
Statistical Consultant 1969-1973

Arthur Funke, Ph.D.
Dissemination Consultant 1969-1974

Stephen Greenwald, M.D.
Medications Consultant 1969-1973

Ann Russell
Clinical Consultant 1970

Robert Spano, A.C.S.W.
Patient Follow-up Coordinator 1969-1973

Wyman Spano
Editorial Consultant 1972-1973

Robert Walker, M.A.
Data Applications Coordinator 1973

David J. Weiss, Ph.D.
Special Statistical Reviewer 1974